DOCUMENT RESUME

ED 179 597 TM 009 986

AUTHOR Garcia-Quintana, Roan A.; Johnson, Lynne M.
TITLE Equating Two Forms of a Criterion-Referenced Test by Using Norm Referenced Data: An Illustration of Two Methods.
INSTITUTION CTB / McGraw-Hill, Monterey, Calif.
PUB DATE Apr 79
NOTE 36p.; Paper presented at the Annual Meeting of the American Educational Research Association (63rd, San Francisco, CA, April 8-12, 1979)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Achievement Tests; Calculation; *Comparative Statistics; Cutting Scores; *Equated Scores; Intermediate Grades; *Mastery Tests; Mathematical Formulas; Norms; Statistical Analysis
IDENTIFIERS *Angoff Method; Quantitative Tests; *Test Equivalence

ABSTRACT

Three different computational procedures for equating two forms of a test were applied to a pair of mathematics tests to compare the results of the three procedures. The tests that were being equated were two forms of the SRA Mastery Mathematics Tests. The common, linking test used for equating was the Comprehensive Tests of Basic Skills, Form S, Mathematics Test. The equating sample was a group of 2,039 sixth grade students in the South Carolina Statewide Testing Program. The first equating method, developed by Angoff, involved estimating the mean and variance for each SRA test for the total group, that is, combining the groups that had taken each test, and then computing a conversion equation from one score scale to the other. The second method, also developed by Angoff, involved equating each of the two SRA forms to the equating test. The third method involved finding a prediction equation for each SRA test from the common test and then combining the two prediction equations to form the equating equation. (CTM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED179597

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

EQUATING TWO FORMS OF A CRITERION-REFERENCED TEST BY USING NORM REFERENCED DATA: AN ILLUSTRATION OF TWO METHODS

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Roan A. Garcia-Quintana

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Roan A. Garcia-Quintana
Statistician
Office of Research
S. C. Department of Education

Lynne M. Johnson
Education Program Specialist
Office of Research
S. C. Department of Education

Presented at the annual meeting of the
American Educational Research Association
San Francisco, California
April 8 - 12, 1979

IM009 986

INTRODUCTION

Typically, the proposed competency-based systems of education have incorporated the meeting of a certain minimal standard as a requirement for a student's progression from one grade to the next or graduation from high school. The consequences of not meeting these standards are of enough social concern that test security has become a problem which demands the attention of measurement specialists in arriving at viable solutions.

One possible solution is the implementation of complex procedures that account for the distribution and collection of all tests administered However, such procedures are costly in terms of time, money, and personnel. Furthermore, these expenditures may not be commensurate with the extent to which the test is secure.

Another possible solution is the use of parallel forms of the same test. By using truly parallel forms, the same minimum standard used for one form would also serve as the minimum standard on the parallel form. However, the establishment of parallel forms in itself is a problem. The necessary (but not sufficient) conditions for demonstrating that two forms are parallel are:

1. the average score on Form X equals the average score on Form Y;
2. the variance of Form X equals the variance of Form Y; and
3. the correlation between the error component on each form is 0 (Lord and Novick, 1968).

These conditions assume that both forms of the test are administered to

independent random samples, where the primary sampling unit is the student. This requirement of randomness is rarely attainable because of practical and logistical problems associated with administering tests within the public school setting.

A third possible solution is the use of alternate forms of the same test and converting the minimum standard on one form to its equivalent score on the other form. By using such an approach, the minimum standard determined for one form [whether on the basis of the number of items correct (raw score) or on the basis of the number of objectives mastered] would be equated to its corresponding equivalent score on the alternate form. Given the practical limitations of using the same "secure" test and the measurement problems involved with constructing two "truly" parallel forms, this third solution appears to be the most feasible from both a practical and statistical standpoint.

Two equating methods developed by Angoff (1971) are applicable to the situation where (1) each of two forms of a given test is administered to each of two groups of examinees, (2) one common equating test is administered to both groups, and (3) the two groups are selected using nonrandom procedures or using random procedures where the individual student is not the primary sampling unit. The group selection feature is especially appealing to practitioners in the public educational setting where exercising freedom in the selection of student groups is not usually possible, resulting in nonrandom groups rather than random groups. Furthermore, where random group selection is attempted in the school setting, the sampling unit is typically some designated group of students rather than the individual student.

The South Carolina Legislation under the 1977 School Finance Act stated that "the State Department of Education shall develop or select and field

test a competency-based student assessment program in the basic skill areas of reading and mathematics utilizing criterion-referenced tests". To comply with the legislation, two different forms of a criterion-referenced test were administered to two different nonrandom groups of sixth graders. In the South Carolina Statewide Testing Program, a norm-referenced test was administered to both groups.

Thus, the South Carolina situation with respect to both test administration and examinee groups selection met the general parameters of the conditions necessary for illustrating the two equating methods by Angoff. In this paper, the applications of these two methods to the South Carolina situation are demonstrated. SRA data for two different forms and CTBS/S data were available for all sixth graders who participated in the field test project.

SAMPLE

In the field test project, two forms of the SRA Mastery Mathematics Test, Forms X and Y (the criterion-referenced measures), were administered to two groups of sixth grade students in Spring of 1978. Within the month following the SRA field test administration, the Comprehensive Tests of Basic Skills, Form S (CTBS/S) (the norm-referenced measure), was administered to all sixth grade students in the state.

Matched SRA Mastery Mathematics Test and CTBS/S Mathematics Test data were available for a total 2039 sixth grade students. Of this group of students, 1257 students had matched data for SRA Mastery Mathematics Test, Form X, and the CTBS/S Mathematics Test; and 782 students had matched data for the SRA Mastery Mathematics Test, Form Y, and the CTBS/S Mathematics Test.

The original field test project sample of 3050 students was selected

utilizing a stratified sampling procedure to ensure a representative statewide sample in terms of achievement. The four achievement strata utilized were based on the percentages of South Carolina sixth grade students scoring in each of four national quarters of the Spring, 1978, CTBS/S Battery Total distribution. Schools and then classes were randomly sampled. However, in some cases, practical limitations prevented the use of random selection techniques. Within each district, schools were, for the most part, randomly assigned either Form X or Form Y of the SRA test. However, in order for the number of sixth grade students taking each form to be approximately the same, a particular test form was assigned deliberately where necessary.

The ideal situation would have been to have had random groups where primary sampling unit was the individual student. However, in the given situation, although random techniques were initially envisioned for the selection of the field test sample and the assignment of the two SRA test forms, the resulting two groups of sixth graders were, nevertheless, nonrandom. As already pointed out, practical limitations prevented two truly random groups of sixth grade students from being selected for the administration of the two criterion-referenced test forms. An additional restriction which resulted in nonrandom groups was the necessity of having matched data. Matched criterion-referenced and norm-referenced data were not available for all students tested as part of the field test sample. The rate of occurrence of the "unmatched" cases was certainly not random in the traditional sense. In addition to the two groups of students being nonrandom, the sampling procedure for the situation of interest deviated from the ideal situation in another way: the primary samplimg unit was the school rather than the individual student.

Thus, there existed two nonramdom groups of sixth graders, with each

group having been administered a different form of the SRA Mastery Mathematics Test and with both groups having been administered the CTBS/S Mathematics Test as the common test.

PROCEDURES

As pointed out by Angoff (1971), in equating, the units of one test form are converted to the units of another test form, that is, another form of the same test. This implies two requirements: (1) that the two test forms being equated measure the same characteristic, and (2) that the resulting conversion is independent of the individuals on which the conversion was based. Hence, the derivation of equivalent scores for two different test forms can only be accomplished if the two forms measure the same psychological trait. The two forms of the SRA Mastery Mathematics Test were constructed to be similar in content. Both Form X and Form Y of the SRA test measure the same forty (40) mathematics objectives with three (3) items measuring each objective.

In addition to the consideration of the content measured, the difficulty of the two test forms being equated is also a consideration. Equivalent scores are usually defined as follows: "Two scores, one on Form X and the other on Form Y (where X and Y measure the same function with the same degree of reliability), may be considered equivalent if their corresponding percentile ranks in any given group are equal" (Angoff, 1971, p. 563). Equating methods based on this definition (usually referred to as equipercentile of curvilinear methods) account for differences in test difficulty by stretching and compressing the raw score distributions of the two test forms so that they coincide. However, if differences in test difficulty are not a problem, (that is, if the shapes of the raw score distributions are the same), linear equating methods are acceptable.

In linear methods, the scores which correspond to equal standard score deviates are equivalent. Therefore, the raw score distributions of the two SRA test forms were first examined to determine whether to use a curvilinear or simply a linear equating method. As illustrated in Figures 1 and 2, the shapes of the cumulative percentage distributions for Forms X and Y of the SRA test were very similar. Consequently, linear equating methods were decided upon.

The situation of interest is: two different criterion-referenced test forms were administered to two different nonrandom groups of examinees and one common norm-referenced equating test was administered to both groups. The following two methods of equating test scores are applicable to this situation:

Method I: Nonrandom groups where one test is administered to each of two groups and a common equating test is administered to both groups to adjust for differences between the two groups.

Method II: Random or nonrandom groups where one test is administered to each of two groups and a common equating test is administered to both groups.

Method II.A: Each of two test forms equated to a common test.

Method II.B: Each of two test forms predicted by a common test.

Four general criteria must hold in order to apply these two equating methods.

Criterion 1: <u>The two test forms being equated measure the same characteristic</u>. In the given situation, SRA Form X and Form Y both measure the same forty (40) mathematics objectives, three (3) items per objective, where the items are very similar in terms of content.

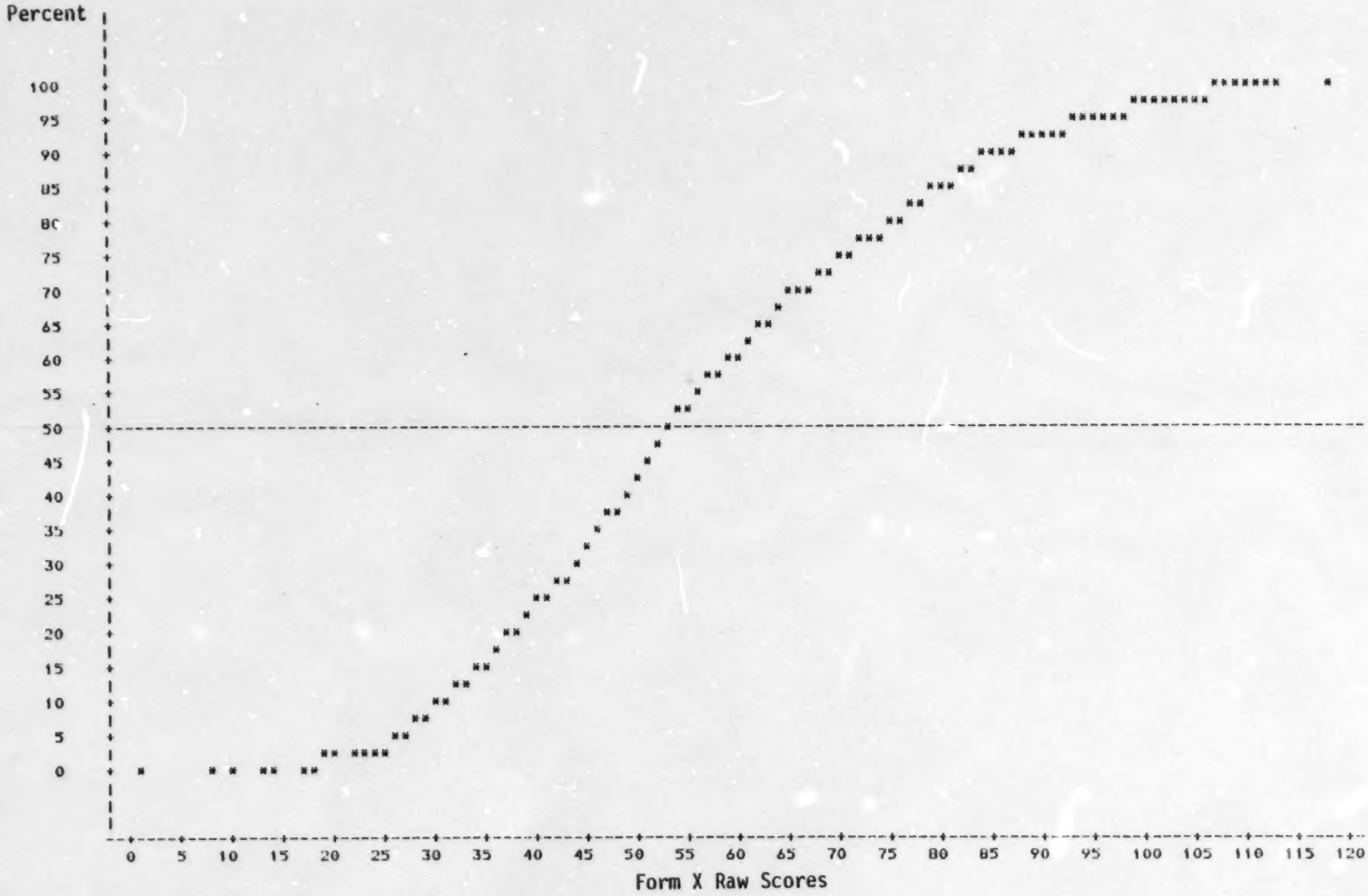

Figure 1

Cumulative Percentage Distribution of SRA Form X

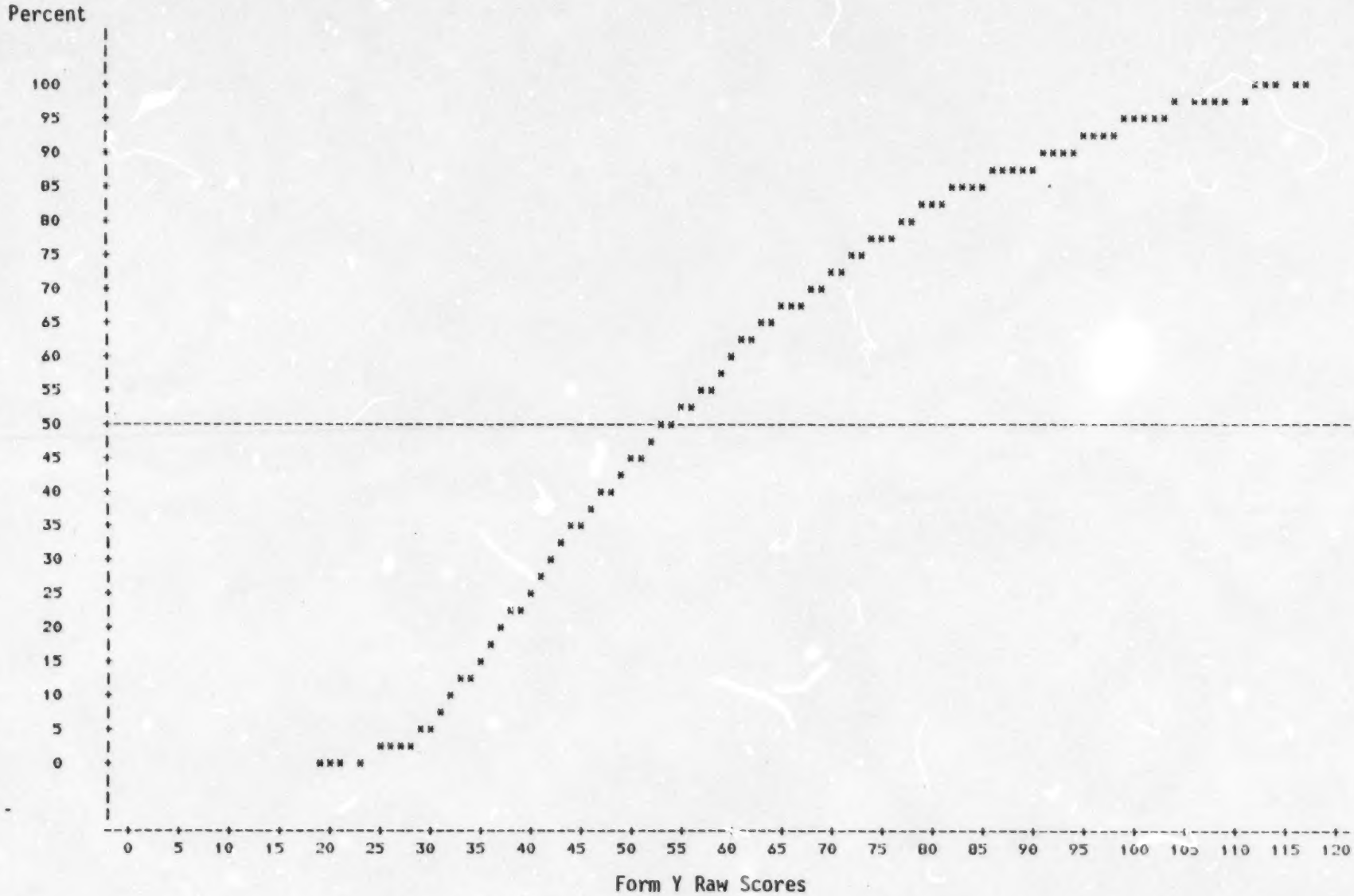

Figure 2

Cumulative Percentage Distribution of SRA Form Y

Criterion 2: The common equating test measures the same characteristic as the two forms being equated. In the given situation, the Mathematics Test of the CTBS/S, Level 2, was analyzed by South Carolina Department of Education consultants as measuring mathematics skills similar to those measured by the two SRA test forms.

Criterion 3: The two groups do not differ widely in ability. In the given situation, the two groups of sixth graders were assigned randomly by school for the most part. Also, the combined group of sixth graders was selected on the basis of their school's prior overall achievement. By virtue of the selection and assignment procedures, the two groups should have been fairly similar in overall achievement. The means for the two groups of sixth graders on Forms X and Y were 56.7438 and 58.3798, respectively, which suggested that these two groups did not differ widely on mathematics achievement.

Criterion 4: The shapes of the raw score distributions of the two test forms being equated are the same. As already pointed out, examination of Figures 1 and 2 suggested that the cumulative percentage distributions for the two different SRA forms are very similar in shape. The distributions differed the most at the tails.

Violations of these criteria and possible alternative approaches are discussed individually with reference to the two methods illustated.

The remainder of the discussion will focus upon the illustration of the two equating methods, with the second method consisting of two parts. Although several types of scores were available for each of the tests administered (for example, scale scores, raw scores, and percentile ranks), the illustrations of the two methods are limited to the use of the raw scores for all tests invo.ved.

For purposes of clarity, the SRA Mastery Mathematics Tests, Forms X and Y, are referred to simply as Form X and Form Y, respectively. The group of students to whom Form X was administered is referred to as Group a, and the group of students to whom Form Y was administered is referred to as Group b. The CTBS/S Mathematics Test, which is the common equating test administered to both groups, is referred to as Test U. The combined group of students (Group a + Group b) is referred to as Group t.

ILLUSTRATION OF METHOD I

Method I is the application of the basic linear equating method for groups not widely different in ability, referred to by Angoff (1971) in Design IV. This method is appropriate for nonrandom groups with two different forms of the same test administered to each of two groups and with a common equating Test U administered to both groups. Although the two groups may have been selected to ensure similarity with respect to certain characteristics, differences may still exist between the two groups. The common equating Test U provides a means of observing the differences that do exist between the two groups and for making adjustments for these differences. Test U can be administered in one of several ways:

1. in addition to, but separate from, Form X and Form Y;
2. as part of Form X and Form Y; and
3. separate from Form X but as part of Form Y, that is, Form Y would have to be administered to both groups (Angoff, 1971).

Regardless of the manner in which Test U is administered and constructed, Test U, nevertheless, must "represent psychologically the same task to both groups" (Angoff, 1971, p. 580).

In Method I, scores on the SRA Mastery Mathematics Tests, Forms X and Y, are equated after adjustments for differences in mathematics achievement are made on the basis of the CTBS/S Mathematics Test, the common norm-referenced equating test. The two groups to whom the criterion-referenced tests were administered were considered to be nonrandom and to be fairly similar with respect to mathematics achievement. Furthermore, the CTBS/S Mathematics Test was analyzed as measuring mathematics skills similar to those measured by the

two SRA test forms. Also recall that the CTBS/S was administered in addition to the SRA tests on a separate occasion following the SRA test administration. Therefore, the situation of interest met the criteria of Angoff's Design IV.

Linear methods for groups widely different in ability are available if Criterion 3 in the Procedures Section does not hold. In the event that Criterion 4 in the Procedures Section is violated, the curvilinear analog for groups not widely different in ability is applicable (Angoff, 1971).

The following procedure illustrates the application of Method I to the situation of interest.

Step 1: For the combined group of examinees, Group t = Group a + Group b, compute:

(a) Mean of common norm-referenced Test U denoted M_{u_t}.

$M_{u_t} = 63.71260$

(b) Variance of common norm-referenced Test U, denoted $S^2_{u_t}$.

$S^2_{u_t} = 354.61216$

Step 2: For each of Groups a and b, individually compute:

(a) Mean of common norm-referenced Test U, denoted M_{u_a} for Group a, and M_{u_b} for Group b.

$M_{u_a} = 63.67462$

$M_{u_b} = 63.77366$

(b) Variance of common norm-referenced Test U, denoted $S^2_{u_a}$ for Group a, and $S^2_{u_b}$ for Group b.

$S^2_{u_a} = 332.76900$

$S^2_{u_b} = 390.18814$

(c) Means of criterion-referenced test Forms X and Y (administered to Groups a and b, respectively), denoted M_{x_a} for Group a and M_{y_b} for Group b.

$M_{x_a} = 56.74383$

$M_{y_b} = 58.37980$

(d) Variances of criterion-referenced test Forms X and Y (administered to Groups a and b, respectively), denoted $S^2_{x_a}$ for Group a and $S^2_{y_b}$ for Group b.

$S^2_{x_a} = 438.90885$

$S^2_{y_b} = 493.17823$

Step 3: Regress each form of the criterion-referenced test, Form X and Form Y, on the norm-referenced Test U. In other words, predict each of the SRA Mastery Forms X and Y from the CTBS/S. Denote the regression coefficient of Form X on Test U for Group a as b_{xu_a}. Denote the regression coefficient of Form Y on Test U for Group b as b_{yu_b}.

$b_{xu_a} = .86458$

$b_{yu_b} = .80370$

Step 4: For the combined Group t composed of examinees in Group a and Group b, compute:

(a) Estimate of the mean for Group t on Form X, denoted $\hat{M}_{x_t}$ where $\hat{M}_{x_t} = M_{x_a} + b_{xu_a}(M_{u_t} - M_{u_a})$.

$\hat{M}_{x_t} = 56.74383 + .86458\ (63.7126 - 63.67462)$

$= 56.77667$

(b) Estimate of the variance for Group t on Form X, denoted $\hat{S}^2_{x_t}$ where $\hat{S}^2_{x_t} = S^2_{x_a} + b_{xu_a}(S^2_{u_t} - S^2_{u_a})$.

$\hat{S}^2_{x_t} = 438.90885 + .86458\ (354.61216 - 332.7690)$

$= 505.81559$

(c) Estimate of the mean for Group t on Form Y, denoted $\hat{M}_{y_t}$ where $\hat{M}_{y_t} = M_{y_b} + b_{yu_b}(M_{u_t} - M_{u_b})$.

$\hat{M}_{y_t} = 58.3798 + .9037\ (63.7126 - 63.77366)$

$= 58.32462$

(d) Estimate of the variance for Group t on Form Y, denoted $\hat{S}^2_{y_t}$ where $\hat{S}^2_{y_t} = S^2_{y_b} + b_{yu_b}(S^2_{u_t} - S^2_{u_b})$.

$\hat{S}^2_{y_t} = 493.17823 + .90370\ (354.61216 - 393.18814)$

$= 461.02822$

Step 5: Substitute the appropriate values to determine the following slopes and intercepts:

(a) Slope of the equation for converting scores on Form X to scores on Form Y, denoted $A_y = \hat{S}_{y_t}/\hat{S}_{x_t}$.

$A_y = 21.47157/22.49034$

$= .95470$

(b) Slope of the equation for converting scores on Form Y to scores on Form X, denoted $A_x = \hat{S}_{x_t}/\hat{S}_{y_t}$.

$A_x = 22.49034/21.47157$

$= 1.04745$

(c) Intercept of the equation for converting scores on Form X to scores on Form Y, denoted $B_y = \hat{M}_{y_t} - A_y \hat{M}_{x_t}$.

$B_y = 58.32462 - .95470\ (56.77667)$

$= 4.11993$

(d) Intercept of the equation for converting scores on Form Y to scores on Form X, denoted $B_x = \hat{M}_{x_t} - A_x \hat{M}_{y_t}$.

$B_x = 56.77667 - 1.04745(58.32462)$

$= -4.31545$

Step 6: By substituting the appropriate values, determine the following conversion equations.

(a) Equation for converting scores on criterion-referenced Form X to scores on criterion-referenced Form Y, given by $Y = A_yX + B_y$

$Y = (.95470)\ X + 4.11993$

(b) Equation for converting scores on criterion-referenced Form Y to scores on criterion-referenced Form X, given by $X = A_X Y + B_X$

$X = (1.04745)Y - 4.31545$

ILLUSTRATION OF METHOD II

Method II is the application of the design referred to by Angoff (1971) as Design V. This method for equating scores on two different forms of a test requires the administration of two different forms (Forms X and Y) of a test to two different groups of examinees (Groups a and b) and the administration of a common test (Test U) to both groups of examinees (Group t). This common test should be administered either preceding or following the administration of Form X and Form Y. The underlying assumption is that, if Forms X and Y are alternate forms of the same test, Test U has the same practice effect on Forms X and Y, regardless of whether the two forms are administered before or after Test U. In the given situation, the criterion-referenced SRA Mastery Mathematics Test, Forms X and Y, were constructed as alternate forms, and the norm-referenced CTBS/S was administered to the combined group of sixth graders after the administration of the SRA tests.

According to Angoff (1971), the additional common test must measure the same content as the other two test forms if there is to be flexibility in the selection of Groups a and b. Recall that the CTBS/S Mathematics Test was analyzed as measuring mathematics skills similar to those measured by the SRA Mastery Mathematics Tests, Forms X and Y. Furthermore, the two groups of examinees to whom SRA Forms X and Y were administered, which do not differ widely in mathematics ability, are considered to be nonrandom groups. Therefore, the situation of interest met the criteria of Angoff's Design V.

In this section, two methods of equating, rather than a single method, are actually illustrated. The two methods are classified by Angoff under Design V because the test administration situation and group selection procedures are applicable to both methods. In the event that Criterion 4 in the Procedures Section is not met by a particular situation, the curvilinear analogs of both Methods II.A and II.B are applicable (Angoff, 1971).

Illustration of Method II.A

Method II.A is a linear equating method whereby two test forms are first equated to a common test. A score on Form X is equivalent to a score on Form Y if the two scores are equivalent to the same score on the common test administered to both groups, Test U. Therefore, in the given situation, a score on the SRA Mastery Mathematics Test, Form X, is equivalent to a score on the SRA Mastery Mathematics Test, Form Y, if the two scores are equivalent to the same score on the norm-referenced CTBS/S Mathematics Test. The following procedure illustrates the application of Method II.A to the situation of interest.

Step 1: Refer to Step 2 of Method I and compute M_{u_a}, M_{u_b}, S_{u_a}, S_{u_b}, M_{x_a}, M_{y_b}, S_{x_a}, and S_{y_b}.

$M_{u_a} = 63.67462$ $S_{u_a} = 18.24196$

$M_{u_b} = 63.77366$ $S_{u_b} = 19.75318$

$M_{x_a} = 56.74383$ $S_{x_a} = 20.95015$

$M_{y_b} = 58.37980$ $S_{y_b} = 22.20762$

Step 2: Substitute the appropriate values to determine the following slopes and intercepts:

(a) Slope of the equation for converting scores on the norm-referenced Test U to scores on the criterion-referenced Form X, denoted $A_{xu} = S_{x_a}/S_{u_a}$.

$A_{xu} = 20.95015/18.24196$

$= 1.14846$

(b) Slope of the equation for converting scores on the norm-referenced Test U to scores on the criterion-referenced Form Y, denoted $A_{yu} = S_{y_b}/S_{u_b}$.

$A_{yu} = 22.20762/19.75318$

$= 1.12426$

(c) Intercept of the equation for converting scores on the norm-referenced Test U to scores on the criterion-referenced Form X, denoted $B_{xu} = M_{x_a} - A_{xu} M_{u_a}$.

$B_{xu} = 56.74383 - (1.14846)(63.67462)$

$= -16.38392$

(d) Intercept of the equation for converting scores on the norm-referenced Test U to scores on the criterion-referenced Form Y, denoted $B_{yu} = M_{y_b} - A_{yu} M_{u_b}$

$B_{yu} = 58.37980 - (1.12426)(63.77366)$

$= -13.31838$

Step 3: The equation for converting scores on the norm-referenced Test U to scores on the criterion-referenced Form X is $X = A_{xu}U + B_{xu}$ where X is a score on Form X and U is a score on Test U. This implies that a score on Form X is equivalent to a score on Test U ($X \equiv U$). The equation for converting scores on the norm-referenced Test U to scores on the criterion-referenced Form Y is $Y = A_{yu}U + B_{yu}$ where Y is a score on

Form Y and U is a score on Test U. This implies that a score on Form Y is equivalent to a score on Test U ($Y \equiv U$). Therefore, scores on X and Y equivalent to the same score on Test U are themselves taken to be equivalent ($X \equiv Y$). By solving these two equations for U, two equations which permit direct conversions of scores on Form X to scores on Form Y and vice-versa, result.

(a) Substitute the appropriate values to determine the equation for directly converting scores on criterion-referenced Form X to scores on criterion-referenced Form Y, given by

$$Y = (A_{yu}/A_{xu})X + B_{yu} - (A_{yu}/A_{xu})B_{xu}$$

$$Y = (1.12426/1.14846)X + (-13.31838) + (1.12426/1.14846)(16.38392$$

hence,

$$Y = (.97893)X + 2.72030$$

(b) Substitute the appropriate values to determine the equation for directly converting scores on criterion-referenced Form Y to scores on criterion-referenced Form X, given by

$$X = (A_{xu}/A_{yu})Y + B_{xu} - (A_{xu}/A_{yu})B_{yu}$$

$$X = (1.14846/1.12426)Y + (-16.38392) + (1.14846/1.12426)(13.31838$$

hence,

$$X = (1.02152)Y - 2.77886$$

Illustration of Method II.B

Method II.B is a linear equating method whereby two test forms are first predicted by a common test. A score on Form X is equivalent to a score on Form Y if the two scores are predicted by the same score on the common test administered to both groups, Test U. Therefore, in the given situation, a score on the SRA Mastery Mathematics Test, Form X, is equivalent to a score on the SRA Mastery Mathematics Test, Form Y, if the two scores

are predicted by the same score on the norm-referenced CTBS/S Mathematics Test. The following procedure illustrates the application of Method II.B to the situation of interest.

Step 1: Refer to Steps 1 and 2 of Method I and compute the following means:

$M_{u_a} = 63.67462$

$M_{u_b} = 63.77366$

$M_{x_a} = 56.74383$

$M_{y_b} = 58.37980$

Step 2: Regress the common norm-referenced Test U, on each of the criterion-referenced Forms X and Y. In other words, predict the CTBS/S from each of the SRA test Forms X and Y. Denote the regression coefficient of Test U on Form X as b_{ux_a}. Denote the regression coefficient of Test U on Form Y as b_{uy_b}.

$b_{ux_a} = .65551$

$b_{uy_b} = .71498$

Step 3: The equation for converting scores on the criterion-referenced Form X to scores on norm-referenced Test U is

$$U_x = b_{ux_a} X + D_{ux_a} \qquad \text{(II.B.1)}$$

where U_x is the score on Test U predicted by Form X. X is a score on Form X, and D_{ux_a} is the intercept. This implies that a score on Form X predicts a score on Test U ($X \rightarrow U_x$). By substituting the appropriate values, compute

$D_{ux_a} = M_{u_a} - b_{ux_a} M_{x_a}$.

$D_{ux_a} = 63.67462 - .65551\ (56.74383)$

$= 26.47847$

The equation for converting scores on criterion-referenced Form Y to scores on norm-referenced Test U is

$$U_y = b_{uy_b} Y + D_{uy_b} \qquad \text{(II.B.2)}$$

where U_y is the score on Test U predicted by Form Y, Y is a score on Form Y, and D_{uy_b} is the intercept. This implies that a score on Form Y predicts a score on Test U ($Y \cong U_y$). By substituting the appropriate values, compute $D_{uy_b} = M_{u_b} - b_{uy_b} M_{y_b}$.

$D_{uy_b} = 63.77366 - .71498\ (58.37980)$

$= 22.03327$

Therefore, when the scores predicted by Forms X and Y are equal ($U_x \cong U_y$), then a score on Form X is equivalent to a score on Form Y ($X \cong Y$).

Step 4: By setting Equations II.B.1 and II.B.2 equal to each other and solving for X or Y, two equations which permit direct conversions of scores on Form X to scores on Form Y and vice versa result.

(a) Substitute the appropriate values to determine the equation for directly converting scores on criterion-referenced Form X to scores on criterion-referenced Form Y, given by $Y = (b_{ux_a}/b_{uy_b})X + (D_{ux_a} - D_{uy_b})/b_{uy_b}$ where X is a score on Form X and Y is the equivalent score on Form Y.

$Y = (.65551/.71498)X + (26.47847 - 22.03327)/.71498$

$= (.91682)X + 6.21724$

(b) Substitute the appropriate values to determine the equation for directly converting scores on criterion-referenced Form Y to scores on criterion-referenced Form X, given by $X = (b_{uy_b}/b_{ux_a})Y + (D_{uy_b} - D_{ux_a})/b_{ux_a}$ where Y is a score on Form Y and X is the equivalent score on Form X.

$X = (.71498/.65551)Y + (22.03327 - 26.47847)/.65551$

$= (1.09072)Y - 6.78128$

RESULTS

In sum, six (6) conversion equations were derived by using "three" different equating methods. The following equations were formulated for converting scores on Form X to scores on Form Y:

Method I: $Y = (.95470)X + 4.11993$

Method II.A: $Y = (.97893)X + 2.72030$

Method II.B: $Y = (.91682)X + 6.21724$

In Table 1, the score conversions based on each of these three equations are presented. In Table 2, the mean differences of the scores and the standard deviations of the differences are provided for the "three" methods.

The following equations were formulated for converting scores on SRA Form Y to scores on SRA Form X:

Method I: $X = (1.04745)Y - 4.31545$

Method II.A: $X = (1.02152)Y - 2.77886$

Method II.B: $X = (1.09072)Y - 6.78128$

In Table 3, the score conversions based on each of these three equations are presented. In Table 4, the mean differences of the scores and the standard deviations of the differences are provided for the "three" methods.

Table 1

Scores on Form Y Equated from Scores on Form X

Form X	Form Y		
	METHOD I	METHOD II.A	METHOD II.B
1.00000	5.07463	3.69923	7.13406
2.00000	6.02933	4.67816	8.05083
3.00000	6.98403	5.65709	8.96770
4.00000	7.93873	6.63602	9.88452
5.00000	8.89343	7.61495	10.80134
6.00000	9.84813	8.59388	11.71816
7.00000	10.80283	9.57281	12.63498
8.00000	11.75753	10.55174	13.55180
9.00000	12.71223	11.53067	14.46862
10.00000	13.66693	12.50960	15.38544
11.00000	14.62163	13.48853	16.30226
12.00000	15.57633	14.46746	17.21908
13.00000	16.53103	15.44639	18.13590
14.00000	17.48573	16.42532	19.05272
15.00000	18.44043	17.40425	19.96954
16.00000	19.39513	18.38318	20.88636
17.00000	20.34983	19.36211	21.80318
18.00000	21.30453	20.34104	22.72000
19.00000	22.25923	21.31997	23.63682
20.00000	23.21393	22.29890	24.55364
21.00000	24.16863	23.27783	25.47046
22.00000	25.12333	24.25676	26.38728
23.00000	26.07803	25.23569	27.30410
24.00000	27.03273	26.21462	28.22092
25.00000	27.98743	27.19355	29.13774
26.00000	28.94213	28.17248	30.05456
27.00000	29.89683	29.15141	30.97138
28.00000	30.85153	30.13034	31.88820
29.00000	31.80623	31.10927	32.80502
30.00000	32.76093	32.08820	33.72184
31.00000	33.71563	33.06713	34.63866
32.00000	34.67033	34.04606	35.55548
33.00000	35.62503	35.02499	36.47230
34.00000	36.57973	36.00392	37.38912
35.00000	37.53443	36.98285	38.30594
36.00000	38.48913	37.96178	39.22276
37.00000	39.44383	38.94071	40.13958
38.00000	40.39853	39.91964	41.05640
39.00000	41.35323	40.89857	41.97322
40.00000	42.30793	41.87750	42.89004
41.00000	43.26263	42.85643	43.80686
42.00000	44.21733	43.83536	44.72368
43.00000	45.17203	44.81429	45.64050
44.00000	46.12673	45.79322	46.55732
45.00000	47.08143	46.77215	47.47414
46.00000	48.03613	47.75108	48.39096
47.00000	48.99083	48.73001	49.30778
48.00000	49.94553	49.70894	50.22460
49.00000	50.90023	50.68787	51.14142
50.00000	51.85493	51.66680	52.05824
51.00000	52.80963	52.64573	52.97506
52.00000	53.76433	53.62466	53.89188
53.00000	54.71903	54.60359	54.80870
54.00000	55.67373	55.58252	55.72552
55.00000	56.62843	56.56145	56.64234
56.00000	57.58313	57.54038	57.55916
57.00000	58.53783	58.51931	58.47598
58.00000	59.49253	59.49824	59.39280
59.00000	60.44723	60.47717	60.30962
60.00000	61.40193	61.45610	61.22644

Table 1 (continued)

Scores on Form Y Equated from Scores on Form X

Form X	Form Y		
	METHOD I	METHOD II.A	METHOD II.B
61.00000	62.35663	62.43503	62.14326
62.00000	63.31133	63.41396	63.06008
63.00000	64.26603	64.39289	63.97690
64.00000	65.22073	65.37182	64.89372
65.00000	66.17543	66.35075	65.81054
66.00000	67.13013	67.32968	66.72736
67.00000	68.08483	68.30861	67.64418
68.00000	69.03953	69.28754	68.56100
69.00000	69.99423	70.26647	69.47782
70.00000	70.94893	71.24540	70.39464
71.00000	71.90363	72.22433	71.31146
72.00000	72.85833	73.20326	72.22828
73.00000	73.81303	74.18219	73.14510
74.00000	74.76773	75.16112	74.06192
75.00000	75.72243	76.14005	74.97874
76.00000	76.67713	77.11898	75.89556
77.00000	77.63183	78.09791	76.81238
78.00000	78.58653	79.07684	77.72920
79.00000	79.54123	80.05577	78.64602
80.00000	80.49593	81.03470	79.56284
81.00000	81.45063	82.01363	80.47966
82.00000	82.40533	82.99256	81.39648
83.00000	83.36003	83.97149	82.31330
84.00000	84.31473	84.95042	83.23012
85.00000	85.26943	85.92935	84.14694
86.00000	86.22413	86.90828	85.06376
87.00000	87.17883	87.88721	85.98058
88.00000	88.13353	88.86614	86.89740
89.00000	89.08823	89.84507	87.81422
90.00000	90.04293	90.82400	88.73104
91.00000	90.99763	91.80293	89.64786
92.00000	91.95233	92.78186	90.56468
93.00000	92.90703	93.76079	91.48150
94.00000	93.86173	94.73972	92.39832
95.00000	94.81643	95.71865	93.31514
96.00000	95.77113	96.69758	94.23196
97.00000	96.72583	97.67651	95.14878
98.00000	97.68053	98.65544	96.06560
99.00000	98.63523	99.63437	96.98242
100.00000	99.58993	100.61330	97.89924
101.00000	100.54463	101.59223	98.81606
102.00000	101.49933	102.57116	99.73288
103.00000	102.45403	103.55009	100.64970
104.00000	103.40873	104.52902	101.56652
105.00000	104.36343	105.5[illegible]	102.48334
106.00000	105.31813	106.4[illegible]88	103.40016
107.00000	106.27283	107.46581	104.31698
108.00000	107.22753	108.44474	105.23380
109.00000	108.18223	109.42367	106.15062
110.00000	109.13693	110.40260	107.06744
111.00000	110.09163	111.38153	107.98426
112.00000	111.04633	112.36046	108.90108
113.00000	112.00103	113.33939	109.81790
114.00000	112.95573	114.31832	110.73472
115.00000	113.91043	115.29725	111.65154
116.00000	114.86513	116.27618	112.56836
117.00000	115.81983	117.25511	113.48518
118.00000	116.77452	118.23404	114.40200
119.00000	117.72923	119.21297	115.31882
120.00000	118.68393	120.19190	116.23564

Table 2

Descriptive Statistics for the Differences Between the Equated Y Scores Produced by the Three Methods

Difference	Mean	Variance	Standard Deviation
(Method I - Method II.A)	-0.06628	0.71038	0.84284
(Method I - Method II.B)	0.19443	1.73622	1.31766
(Method II.A - Method II.B)	0.26072	4.66776	2.16050

Table 3

Scores on Form X Equated from Scores on Form Y

Form Y	Form X		
	METHOD I	METHOD II.A	METHOD II.B
1.00000	-3.07248	-1.75734	-5.69056
2.00000	-2.02503	-0.73582	-4.59984
3.00000	-0.97759	0.28570	-3.50912
4.00000	0.06987	1.30722	-2.41840
5.00000	1.11732	2.32874	-1.32768
6.00000	2.16477	3.35026	-0.23696
7.00000	3.21222	4.37178	0.85376
8.00000	4.25967	5.39330	1.94448
9.00000	5.30712	6.41482	3.03520
10.00000	6.35457	7.43634	4.12592
11.00000	7.40202	8.45786	5.21664
12.00000	8.44947	9.47938	6.30736
13.00000	9.49692	10.50090	7.39808
14.00000	10.54437	11.52242	8.48880
15.00000	11.59182	12.54394	9.57952
16.00000	12.63927	13.56546	10.67024
17.00000	13.68672	14.58698	11.76096
18.00000	14.73417	15.60850	12.85168
19.00000	15.78162	16.63002	13.94240
20.00000	16.82907	17.65154	15.03312
21.00000	17.87652	18.67306	16.12384
22.00000	18.92397	19.69458	17.21456
23.00000	19.97142	20.71610	18.30528
24.00000	21.01887	21.73762	19.39600
25.00000	22.06632	22.75914	20.48672
26.00000	23.11377	23.78066	21.57744
27.00000	24.16122	24.80218	22.66816
28.00000	25.20867	25.82370	23.75888
29.00000	26.25612	26.84522	24.84960
30.00000	27.30357	27.86674	25.94032
31.00000	28.35102	28.88826	27.03104
32.00000	29.39847	29.90978	28.12176
33.00000	30.44592	30.93130	29.21248
34.00000	31.49337	31.95282	30.30320
35.00000	32.54082	32.97434	31.39392
36.00000	33.58827	33.99586	32.48464
37.00000	34.63572	35.01738	33.57536
38.00000	35.68317	36.03890	34.66608
39.00000	36.73062	37.06042	35.75680
40.00000	37.77807	38.08194	36.84752
41.00000	38.82552	39.10346	37.93824
42.00000	39.87297	40.12498	39.02896
43.00000	40.92042	41.14650	40.11968
44.00000	41.96787	42.16802	41.21040
45.00000	43.01532	43.18954	42.30112
46.00000	44.06277	44.21106	43.39184
47.00000	45.11022	45.23258	44.48256
48.00000	46.15767	46.25410	45.57328
49.00000	47.20512	47.27562	46.66400
50.00000	48.25257	48.29714	47.75472
51.00000	49.30002	49.31866	48.84544
52.00000	50.34747	50.34018	49.93616
53.00000	51.39492	51.36170	51.02688
54.00000	52.44237	52.38322	52.11760
55.00000	53.48982	53.40474	53.20832
56.00000	54.53727	54.42626	54.29904
57.00000	55.58472	55.44778	55.38976
58.00000	56.63217	56.46930	56.48048
59.00000	57.67962	57.49082	57.57120
60.00000	58.72707	58.51234	58.66192

Table 3 (continued)

Scores on Form X Equated from Scores on Form Y

Form Y	Form X		
	METHOD I	METHOD II.A	METHOD II.B
61.00000	59.77452	59.53386	59.75264
62.00000	60.82197	60.55538	60.84336
63.00000	61.86942	61.57690	61.93408
64.00000	62.91687	62.59842	63.02480
65.00000	63.96432	63.61994	64.11552
66.00000	65.01177	64.64146	65.20624
67.00000	66.05922	65.66298	66.29696
68.00000	67.10667	66.68450	67.38768
69.00000	68.15412	67.70602	68.47840
70.00000	69.20157	68.72754	69.56912
71.00000	70.24902	69.74906	70.65984
72.00000	71.29647	70.77058	71.75056
73.00000	72.34392	71.79210	72.84128
74.00000	73.39137	72.81362	73.93200
75.00000	74.43882	73.83514	75.02272
76.00000	75.48627	74.85666	76.11344
77.00000	76.53372	75.87818	77.20416
78.00000	77.58117	76.89970	78.29488
79.00000	78.62862	77.92122	79.38560
80.00000	79.67607	78.94274	80.47632
81.00000	80.72352	79.96426	81.56704
82.00000	81.77097	80.98578	82.65776
83.00000	82.81842	82.00730	83.74848
84.00000	83.86587	83.02882	84.83920
85.00000	84.91332	84.05034	85.92992
86.00000	85.96077	85.07186	87.02064
87.00000	87.00822	86.09338	88.11136
88.00000	88.05567	87.11490	89.20208
89.00000	89.10312	88.13642	90.29280
90.00000	90.15057	89.15794	91.38352
91.00000	91.19802	90.17946	92.47424
92.00000	92.24547	91.20098	93.56496
93.00000	93.29292	92.22250	94.65568
94.00000	94.34037	93.24402	95.74640
95.00000	95.38782	94.26554	96.83712
96.00000	96.43527	95.28706	97.92784
97.00000	97.48272	96.30858	99.01856
98.00000	98.53017	97.33010	100.10928
99.00000	99.57762	98.35162	101.20000
100.00000	100.62507	99.37314	102.29072
101.00000	101.67252	100.39466	103.38144
102.00000	102.71997	101.41618	104.47216
103.00000	103.76742	102.43770	105.56288
104.00000	104.81487	103.45922	106.65360
105.00000	105.86232	104.48074	107.74432
106.00000	106.90977	105.50226	108.83504
107.00000	107.95722	106.52378	109.92576
108.00000	109.00467	107.54530	111.01648
109.00000	110.05212	108.56682	112.10720
110.00000	111.09957	109.58834	113.19792
111.00000	112.14702	110.60986	114.28864
112.00000	113.19447	111.63138	115.37936
113.00000	114.24192	112.65290	116.47008
114.00000	115.28937	113.67442	117.56080
115.00000	116.33682	114.69594	118.65152
116.00000	117.38427	115.71746	119.74224
117.00000	118.43172	116.73898	120.83296
118.00000	119.47917	117.76050	121.92368
119.00000	120.52662	118.78202	123.01440
120.00000	121.57407	119.80354	124.10512

Table 4

Descriptive Statistics for the Differences Between the Equated X Scores Produced by the Three Methods

Difference	Mean	Variance	Standard Deviation
(Method I - Method II.A)	0.22770	0.81356	0.90110
(Method I - Method II.B)	0.04352	2.26547	1.50515
(Method II.A - Method II.B)	-0.18418	5.79425	2.40713

DISCUSSION

Given two truly parallel forms of a test, with each form having the same number of items, the scores on one form should be equal to the equated scores on the other parallel form. This is not the case for alternate forms as seen in the illustrations of the two methods. With respect to the situation of interest, the average raw score on SRA Form X ($\overline{X}$ = 56.74383) was slightly lower than that on SRA Form Y ($\overline{Y}$ = 58.37980). Hence, in the conversion of scores on Form X to scores on Form Y, equated scores on Form Y that are slightly higher than the actual scores on Form X are expected. Examination of Table 1 indicates that this expectation was indeed the case for each of the three methods used. The same expectation is supported by the scores presented in Table 3 where the scores on SRA Form Y are equated to lower scores on SRA Form X. Table 1 indicates that Method II.A consistently yielded the lowest equated SRA Form Y scores, followed by Method I, and then by Method II.B. As would be expected, this order is reversed for the equated SRA Form X scores in Table 3.

Theoretically, when one form of a test is equated to an alternate form of the same test, if all assumptions are met, the equated scores should not be a function of the method used. The scores yielded by one method should be equal to the scores yielded by any other method. This is not the case in actuality as indicated by the mean differences for across the three methods given in Tables 2 and 4. The mean differences for the Y scores produced by the three methods ranged from -.06628 to .26072, and the mean differences for the X scores produced by the three methods ranged from .04352 to .22770. By rounding to unity the following difference ranges are were obtained. For raw scores in SRA Form X in the range of 1 to 44, the equated scores on SRA Form Y

varied from a minimum of one (1) to a maximum of three (3) across the three methods (refer to Table 1). For raw scores on SRA Form X in the range of 45 to 68, the equated scores on SRA Form Y were essentially the same across the three methods (refer to Table 1). For raw scores on SRA Form X in the range of 69 to 120, the equated scores on SRA Form Y varied from a minimum of one (1) to a maximum of four (4) (refer to Table 1). As indicated in Table 2, the same difference ranges across the three methods existed, with the respective raw score ranges on Form Y being 1 to 48, 49 to 69, and 70 to 120. As expected, the differences across the three methods were most pronounced at the extremes of the score distributions. Note that these differences reflect the differences evidenced in the tails of the raw score distributions of SRA Form X and Form Y (refer to Figures 1 and 2).

RECOMMENDATIONS

The differences among the three equating methods were for the most part a function of the data used. Hence, no recommendation as to the best method is made on the basis of the illustrations of the three methods. However, several recommendations are made with respect to considerations in selecting one of the three methods. All three recommendations require that all three methods be applied to the data.

A first recommendation is that the practitioner select that method for which the data satisfies all or the majority of the criteria as the preferred equating method.

A second recommendation is that the decision with respect to which equating method to use should be a function of educational considerations. For instance, in many cases, the equating method would be used to determine the equated cut-off score on one form of a criterion-referenced test, after a cut-off score on the other form of the test has been decided. The practitioner may prefer to select the method that yields the lowest equated score for the predetermined cut-off. For example, if a cut-off score of 90 on Form X had been predetermined, the practitioner may select as the cut-off score on Form Y the score of 89 yielded by Method II.B. The result would be that fewer students would be failed than if one of the other two methods were selected (refer to Table 1). In other words, certain students who would have otherwise failed would be passed. On the other hand, the practitioner may decide upon an approach with different educational consequences. The practitioner who desires to fail rather than pass more students would select as the cut-off score, the highest converted score on the other form. In the case of the cut-off score of 90 on SRA Form X, the score of 91 yielded by Method II.A would be preferred (refer to Table 1). That is, certain students who would have other-

wise passed, would be failed. This recommendation is not meant to imply that the equating method selected should be changed over time to accommodate the parameters of the educational situation. Upon examination of all three methods, one approach should be selected and followed.

Since the three methods did exhibit differences at the score distribution extremes, the third and final recommendation is made that a range of equated scores be used for a cut-off score which falls in the distribution extremes. For example, suppose a cut-off score of 90 on SRA Form X is decided upon. In setting the cut-off score on Form Y, the recommendation is that any score on Form Y falling in the range of 88 to 91 be considered acceptable. This approach is viewed as ethically the most satisfying.

In those cases where the criteria for these three methods are grossly violated, the practitioner should not use any of the methods illustrated and consider alternative equating methods such as latent trait methods outlined in the literature.

REFERENCES

Angoff, W. H. Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.), Educational Measurement. Washington: American Council on Education, 1971. Pp. 508-600.

Hambleton, R. K. and Cook, L. L. Latent trait models and their use in the analysis of educational test data. Journal of Educational Measurement, V14, n2, 1977.

Lord, F. M. and Novick, M. R. Statistical Theories of Mental Test Scores. Reading, Massachusetts: Addison-Wesley, 1968.

South Carolina Validation Study of Criterion-Referenced Tests in Reading and Mathematics, Office of Research Report Series, South Carolina Department of Education, August, 1978.

South Carolina Code of Laws, 1976, Section 59-20-60 (5)(d).